



Technological University Dublin  
**ARROW@TU Dublin**

---

Conference papers

School of Computing

---

2016-1

## Using Topic Modelling Algorithms for Hierarchical Activity Discovery

Eoin Rogers

*Technological University Dublin, [eoinrogers@gmail.com](mailto:eoinrogers@gmail.com)*

John D. Kelleher

*Technological University Dublin, [john.d.kelleher@tudublin.ie](mailto:john.d.kelleher@tudublin.ie)*

Robert Ross

*Technological University Dublin, [robert.ross@tudublin.ie](mailto:robert.ross@tudublin.ie)*

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomcon>

 Part of the [Artificial Intelligence and Robotics Commons](#)

---

### Recommended Citation

Rogers, E., Kelleher, J. & Ross, R. (2016) Using Topic Modelling Algorithms for Hierarchical Activity Discovery, 7th International Conference on Ambient Intelligence, Seville, Spain, 1-3 June 2016.

This Conference Paper is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact [yvonne.desmond@tudublin.ie](mailto:yvonne.desmond@tudublin.ie), [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [brian.widdis@tudublin.ie](mailto:brian.widdis@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)



# Using Topic Modelling Algorithms for Hierarchical Activity Discovery

Eoin Rogers, John D. Kelleher, and Robert J. Ross

Applied Intelligence Research Centre, Dublin Institute of Technology, Dublin, Ireland

`robert.ross@dit.ie`

`eoin.rogers@student.dit.ie`

`john.d.kelleher@dit.ie`

**Abstract.** Activity discovery is the unsupervised process of discovering patterns in data produced from sensor networks that are monitoring the behaviour of human subjects. Improvements in activity discovery may simplify the training of activity recognition models by enabling the automated annotation of datasets and also the construction of systems that can detect and highlight deviations from normal behaviour. With this in mind, we propose an approach to activity discovery based on topic modelling techniques, and evaluate it on a dataset that mimics complex, interleaved sensor data in the real world. We also propose a means for discovering hierarchies of aggregated activities and discuss a mechanism for visualising the behaviour of such algorithms graphically.

## 1 Introduction

The goal of activity recognition (AR) is to automatically identify the activities of humans using data collected from networks of sensors. The sensors may be embedded in the environment or can be wearable sensors. The dominant approach to AR is to train activity identification models by running machine learning algorithms on annotated datasets [8]. Typically these datasets have to be manually annotated. This can be problematic in terms of the human labour required to create a dataset and in the accuracy of the annotations.

Activity discovery (AD) has been proposed as an alternative to manual annotation. AD involves the unsupervised discovery of activities from raw sensor data. In this paper we present an approach to AD based on topic modelling. The approach can discover activities and also discover activity hierarchies where high-level complex activities are composed of simple activities. The approach is implemented in a system that is tested on a dataset that mimics complex, interleaved sensor data. As a second form of evaluation we propose a graphical visualisation of the output of the system which is designed to provide insight into the hierarchies of activities that the system identifies.

**Paper structure:** In Sect. 2 we review prior work in the area of activity discovery. Then, Sect. 3 describes our approach to AD. Finally, Sect. 4 presents an evaluation of our system and makes some concluding remarks.

## 2 Prior Work on Activity Discovery

AD is concerned with the identification of activities in a dataset independently of human annotation [3]. An AD system generally works by identifying recurring patterns in the sensor data without attempting to understand what activity the pattern signifies (this semantic information has to be added by a human user at a later date). AD is already a well-established field of research and as a result a number of approaches to the task have been proposed and tested.

Cook et al. [3] notes that many pattern mining algorithms have been used without modification, and with some success, for AD. [3] also propose their own algorithm which uses beam search to discover patterns that best compress the data from candidate patterns generated by extending previously discovered patterns. Genetic algorithms are an interesting way of tackling the activity discovery problem. The genetic algorithm proposed in [9], called GAIS, uses comparisons to pre-discovered patterns (called “*models*” by the authors) as a fitness function. This makes this approach well-suited for detected variations of existing patterns, but it is not optimised for detecting novel patterns.

Another approach in the literature to AD is topic modelling, for example as proposed by [6], in which a sliding window over temporal sensor data is used to produce “*documents*” that can be fed into a topic modelling algorithm. For example, given a dataset consisting of sensor readings over a one hour period, the user selects a temporal period of, say, 120 seconds and an increment of, say, 20 seconds. A sliding window is placed over the dataset starting at 0 seconds and ending at 120 seconds. The contents of the window are copied into a document, and the window is incremented by 20 seconds (so it is now covering the period from 20 seconds to 140 seconds). Another document is produced by copying the window’s contents, and again the window is incremented. This process continues until the end of the dataset is reached. This will produce a number of documents (actually  $n - w + 1$  documents, where  $n$  is the length of the dataset and  $w$  is the window length).

The algorithm proposed by [6] is conceptually simple and it is able to detect novel patterns in previously unseen data. Nonetheless, we take the view that there are a number of issues with the approach taken by this paper that could be improved upon. Firstly, as noted by [5], the dataset used did not contain any *interleaving* (in other words, there were no situations where multiple activities occurred at the same time). Since most real-world AD systems will be expected to deal with interleaved data, we feel that this substantially limits the usefulness of the results presented. On top of that, the sensors used in the original papers were wearable sensors attached to the subject’s body. Such sensors can collect detailed information about the user, since they can potentially record every action they carry out. However, many potential users are likely to find them burdensome, since they have to remember to put them on for the system to work correctly, not to mention the potential privacy issues surrounding collecting detailed information about their movements throughout their waking hours. For this reason, we want to see to what extent [6]’s approach extends to activity discovery using simple binary sensors embedded directly in an environment, without collecting

detailed information on people’s movements. Most significantly however, the system built by [6] does not have a means to abstract the activities it discovers. As the authors themselves note, people tend to structure and name activities in a hierarchical fashion. Finding a way to automatically detect and output these hierarchies could make activity discovery systems produce more useful and far more semantically meaningful outputs.

### 3 Our Approach

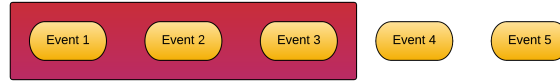
We now present the approach that we developed for this paper. Our approach is based on that presented by [6], and discussed in Sect. 2. It is built on the concept of topic modelling, a concept taken from the natural language processing community. Due to space constraints, we do not include any detailed introduction to topic modelling here, except to say that topic modelling algorithms take a collection of documents as input, and categorise them into a discrete number of topics based on their respective similarities. Like [6], we utilise an approach to topic modelling called Latent Dirichlet Allocation (LDA). We refer the interested reader to [2] for more detail on LDA.

For our analysis and testing we used the *SCARE corpus* [11], a corpus of situated dialogues, where activities were carried out in a virtual environment by teams of two volunteers. One volunteer controlled movement and action in the virtual environment and one directed this controlling volunteer by providing them with instructions on how to complete a list of given tasks. Since the corpus was collected for research into situated dialogues, we had to convert it into a format more suitable for activity discovery. This was done as follows: we produced a list of *events* in the virtual world (events where a door is opened by a user, or where a user picks up an object or places one down) in the temporal order in which they occurred in the original dataset. Note that only the order is preserved: we do not store information about the time at which the events occurred, or even the amount of time that passed between individual events. The reasons for this will be made clear shortly.

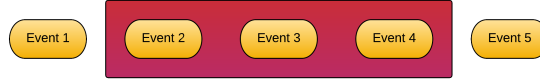
Our specific approach involves the construction of a hierarchy of activities. While this hierarchical process is the primary contribution of this paper, we first discuss the non-hierarchical aspects of the system. Initially, we proceeded to build our system to operate using the same basic algorithm proposed by [6], and discussed in Sect. 2. Figure 1(a) shows an example of a sliding window of length 3 over a dataset of length 5 (in reality, the true SCARE extract we used has a length of over 1600 events). In Fig. 1(b) we can see the sliding window has moved one event forwards. Unlike in the work of [6], we always increment the sliding window by exactly one event at a time, giving our system a larger number of documents to work with. Thus this has produced two documents, and this process continues to generate more documents until the window reaches the end of the dataset. Interestingly, we can also produce a probability vector of topics for each event. This is done by computing the product of the probability distribution vectors for all windows/documents that contain the event, and then

re-normalising the resulting vector to produce a probability vector. Distributions produced by this method will be used in Sect. 4 to evaluate the performance of the system, and they also provide the basis for the hierarchical analysis.

Our hierarchical analysis aims to address the fact that although a pure topic modelling approach provides a somewhat robust means by which activities can be extracted from raw sensor data, it suffers from a substantial problem. It isn't clear what window size should be used when running the system. This is a user-defined parameter, but setting it to different values can result in profound differences in the resulting output. Worse, the resulting differences may not be simply "right" or "wrong", since as the window lengths increase the level of abstraction of the discovered topics will presumably increase. Many activities could plausibly have multiple levels of abstraction, and an activity discovered at any one level could still qualify as correct. For example, if we were processing a dataset that contained events from a kitchen, we might discover an activity at one level of abstraction that corresponded to "making tea". With a smaller window size, we might then find an overlapping activity corresponding to "boiling kettle" (since boiling kettle could be a constituent activity of making tea). Likewise, with a larger window size we might find an overlapping activity called "making dinner" (again, making tea could be a constituent activity of making dinner).



(a) A sliding window covering events 1 to 3



(b) A sliding window covering events 2 to 4



(c) Events 2 and 3 have been abstracted into a single event

**Fig. 1.** An illustration of the internal operation of our approach

In order to make progress in solving this issue, our model follows the principal outlined in Fig. 1(c). Here, *Event 2* and *Event 3* were found with high probability to belong to an activity (the probability distribution vector over topics for that event discussed above has a probability exceeding a user-supplied threshold for at least one activity), and the highest probability activity for both matched.

This means that we remove them both from the dataset, and replace them with a new event, which abstracts away the activity that this topic is presumed to represent. The algorithm is then re-run for a second iteration using this new dataset, allowing for the inference of higher level activities. This means that by using a fairly small, fixed window length, we can allow the system to explore and discover activities over multiple levels of abstraction. A variant of LDA, Hierarchical LDA (hLDA), exists [1] which already allows for hierarchies of topics to be learned. Our system, however, is doing something conceptually different. Whereas hLDA builds a hierarchy of topics based on perceived semantic similarity (i.e. news about football and basketball could be abstracted into a “*sports*” topic), our system builds hierarchies based on constituency, i.e. where one topic (or activity in our case) is a subset of, or overlaps with, another topic.

## 4 Results and Discussion

Evaluating the results of an AD system is a challenge. Typically, due to the relationship between pattern discovery in general and clustering, it is assumed that methods for evaluating clustering algorithms (for instance the Silhouette coefficient and similar metrics) would also be appropriate for pattern discovery and thus for AD. However, most of these metrics work by rewarding the maximisation of dissimilarity between cluster members. In many cases, we can imagine distinct activities having a lot of overlap in terms of what sensors they are likely to activate (both using the toilet and taking a shower will activate sensors in the bathroom, but they are clearly distinct activities). This means that these methods may not be suitable for evaluating the performance of AD systems.

We are, however, using a dataset which contains ground truths, which can be compared to the system’s output. This is also the performance metric utilised by [6]. We used F1 scores as our evaluation metric. In order for this to happen, we had to find a way of matching topics to their most probable corresponding label in the original dataset. This is done via a simple greedy algorithm. Firstly, we extract the most probable topic/activity for each entry in the dataset according to our algorithm. We then iterate over the ground labels. For each ground label, we evaluate the F1 score for each remaining topic, taking the current label to be the corresponding ground truth. We then assign the topic with the highest corresponding F1 score to be the matching topic for our label. The topic is then removed from consideration before the algorithm repeats on the next label (i.e. we will never assign a topic to more than one label). The results of running this evaluation over a window size of 22 (which seems to perform more acceptably than other sizes we have tried) are shown in Table 4.

At first glance, these results might seem to lag behind that of [6]. However, this is due to a number of important features of this work. To begin with, our dataset is fundamentally more difficult in two ways. Firstly, activities are interleaved in a dataset, and so the boundaries between activities is not as clear-cut as in the dataset used for example by [6]. Secondly, we are using binary sensors that are embedded in the environment, rather than on-body sensors producing

**Table 1.** Performance of our system running with window length 22

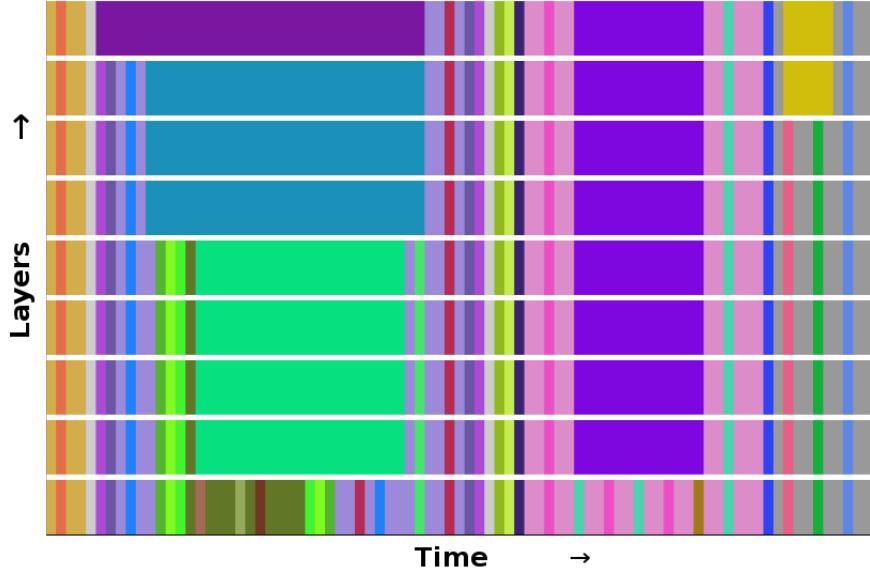
Topic	Label	Precision	Recall	F1 score
Topic 0	Move Quad	0.7353	0.5556	0.6329
Topic 1	Move Picture	0.9780	0.1295	0.2288
Topic 2	Move Rebreather	0.8923	0.4947	0.6365
Topic 3	Move Silencer	0.8015	0.1870	0.3032
Topic 4	Move Box	0.0	0.0	0.0
Topic 5	None	0.0449	0.0270	0.0338

complex motion data. This gives access to low-level motions and actions carried out by the user, producing a considerably richer dataset than in some other cases.

Considering the results in more detail, we see that the *Move Box* label seems to have been particularly difficult for the system to detect. Most instances of this activity are very short, so it may not be a large enough pattern for the system to pick up on. The fifth topic was assigned to a label called *None*. This does not appear in the original SCARE dataset: there are 5 labels taken from the dataset, and a sixth label that we artificially added, which in effect means no activity was taking place.

Another problem that could potentially impact on the performance of AD systems is that there may not be a clean overlap between the natural patterns in the data (which is what we would expect such a system to output) and the hand-annotated activities in the dataset itself, which we are using as ground truths for this evaluation. This indicates that our evaluation metric may in fact not be particularly well suited to our problem. It also provides us with no way of evaluating the hierarchical system that we have built. An alternative solution may simply be to visualise the patterns that have been detected in some way, and then manually inspect them. Figure 2 illustrates an example of a diagram we have produced from the system’s hierarchical output (the complete image is far too wide to reproduce here without loosing an unacceptable degree of detail). This image can be understood as a graph, with time plotted along the x-axis and the layers of the hierarchy along the y-axis. At the bottom of the image, different colours indicate different sensor events occurring over the course of time. As we begin to move up towards the top of the image, we can see sensor events being replaced with new abstract events, corresponding to low-level activities as explained in the previous section. In the centre left of the image, we can see that an abstract event is created at an early stage in the hierarchy, and this event is subsumed into a new, higher level event about half-way up. At the top of the diagram, this new abstraction is in turn subsumed into an even higher one. The very right of the image shows a abstract event being created very late in the hierarchy, but never subsumed by anything. The late formation of abstractions is fairly common, and results from the relative probabilities of events that are carried over from previous layers changing due to the replacement of low-level events with abstract events. On top of this, we can see repeating patterns of

colour in the bottommost layer. These correspond to the activities that the system is producing.



**Fig. 2.** The diagram style we are using to visualise our system’s output

A similarity can be noted between the hierarchical system that we have developed and the concept of maximising compression as the goal of pattern discovery, for example as used by [3]. The concept of using compression in machine learning originates with the *minimum description length* (MDL) principle, as proposed by [7]. The topmost layer in the hierarchy is about 86.56% the length of the full dataset in the bottom layer. This means that we have a mathematical justification for claiming that our system works on the SCARE corpus. This approach should also be applicable to other datasets (and we plan to use it on such in the future), but due to time constraints we could not present results for these in this paper.

In conclusion, we have demonstrated and provided an initial evaluation of an algorithm to hierarchically discover activities using an LDA model. We have also presented a reasonable way to visualise the output of this process. As highlighted previously, interleaving remains a significant challenge for activity discovery systems, so we aim to investigate ways of dealing with it in the future. We also hope to look further into evaluation techniques for AD, including finding ways to determine the degree to which the discovered activities make sense semantically to a human observer. Finally, we aim to utilise our approach on larger and more complex datasets.



## References

1. Blei, D., Griffiths, T., Jordan, M.: The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies. *Journal of the ACM*, Volume 57 Number 2 (2010)
2. Blei, D., Ng, A., Jordan, M.: Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 993–1022 (2003)
3. Cook, D., Krishnan, N., Rashidi, P.: Activity Discovery and Activity Recognition: A New Partnership. *IEEE Transactions on Cybernetics*, Volume 43 Issue 3, 820–828 (2013)
4. Dong, G., Li, J.: Efficient Mining of Emerging Patterns: Discovering Trends and Differences. *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 43–52 (1999)
5. Gu, T., Wang, L., Wu, Z., Tao, X., Lu, J.: A Pattern Mining Approach to Sensor-Based Human Activity Recognition. *IEEE Transactions on Knowledge and Data Engineering*, Volume 23 Number 9, 1359–1377 (2011)
6. Huỳnh, T., Fritz, M., Schiele, B.: Discovery of Activity Patterns using Topic Models. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 10–19 (2008)
7. Rissanen, J.: Modeling by shortest data description. *Automatica* Volume 14 Issue 5, 465–471 (1978)
8. Ross, R., Kelleher, J.: Accuracy and Timeliness in ML Based Activity Recognition. *Proceedings of the AAAI Workshop on Plan, Activity, and Intent Recognition (PAIR)*, Association for the Advancement of Artificial Intelligence, 2013.
9. Ruotsalainen, M., Ala-Keemola, T., Visa, A.: GAIS: A Method for Detecting Interleaved Sequential Patterns from Imperfect Data. *IEEE Symposium on Computational Intelligence and Data Mining, 2007 (CIDM 2007)*, 530–534 (2007)
10. Saives, J., Pianon, C., Faraut, G.: Activity Discovery and Detection of Behavioural Deviations of an Inhabitant From Binary Sensors. *IEEE Transactions on Automation Science and Engineering*, Volume 12 Number 4, 1211–1224 (2015)
11. Stoia, L., Shockley, D., Byron, D., Fosler-Lussier, E.: SCARE: A Situated Corpus with Annotated Referring Expressions. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (2008)